JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

Impact of the Flipped Classroom on Student Performance and Retention: A Parallel Controlled Study in General Chemistry

Michael D. Ryan* and Scott A. Reid*

Department of Chemistry, Marquette University, Milwaukee, Wisconsin 53233, United States

Supporting Information

ABSTRACT: Despite much recent interest in the flipped classroom, quantitative studies are slowly emerging, particularly in the sciences. We report a year-long parallel controlled study of the flipped classroom in a second-term general chemistry course. The flipped course was piloted in the off-semester course in Fall 2014, and the availability of the flipped section in Spring 2015 was broadly advertised prior to registration. Students self-selected into the control and flipped sections, which were taught in parallel by the same instructor; initial populations were 206 in the control section, 117 in the flipped. As a pretest, we used the ACS first-term general chemistry exam (form 2005), given as the final exam across all sections of the first-term course. Analysis of pretest scores, student percentile rankings in the first-term course, and population demographics indicated very similar populations in the two sections. The course designs required comparable student effort, and five common exams were administered, including as a final the ACS second-term general chemistry exam (form 2010). Exam items were validated using classical test theory and Rasch analysis. We find that exam performance in the two sections is statistically different only for the bottom third, as measured by pretest score or percentile rank; here improvement was seen in the flipped class across all five exams. Following this trend was a significant (56%) decrease in DFW percentage (Ds, Fs, withdrawals) in the flipped courses as compared with the control. While both courses incorporated online homework/assessments, the correlation of this indicator with exam performance was stronger in the flipped section, particularly among the bottom demographic. We reflect on the origin and implication of these trends, using data also from student evaluations.

KEYWORDS: *First-Year Undergraduate/General, Hands-On Learning/Manipulatives, Chemical Education Research, Curriculum, Distance Learning/Self Instruction, Internet/Web-Based Learning*

FEATURE: Chemical Education Research

■ INTRODUCTION

Since its inception in the early 2000s,[1−3] the flipped classroom concept has gained in popularity.[4−6] While many models exist, and have been referred to by other names including "inverted teaching",[2,3,7] at its core the flipped concept involves moving classrooms from teacher to student centered, and instructors from lecturers to facilitators.[8] The flipped approach pushes lecture content outside of class, often using short video lectures and quizzes, and the classroom structure bears an increased focus on interactive engagement, e.g., through problem solving and application.[9] Simply stated, the goal of the flipped classroom is to increase student engagement and ownership of their learning.

In assessing the flipped classroom, we first note that much research supports the efficacy of student-centered, active-learning methodologies in facilitating student success. For example, active learning and interactive engagement approaches have been repeatedly shown to increase student academic success[10−15] and lead to other positive outcomes including improved attitudes and fewer misconceptions.[16] While the framework of the flipped approach is thus solid, its impact on student success and retention remains somewhat controversial, as quantitative studies are limited, and peer-reviewed reports in the natural and physical sciences are relatively few.[17,18] As pointed out in reviews of research related to the flipped classroom,[19,20] studies to date have often relied on indirect evidence such as student satisfaction and course grades, and there is little objective evidence based upon controlled experimental designs. While much of the current evidence is indirect, initial results in the sciences and engineering nonetheless have shown promise, with reports of higher scores and lower failure and DFW (Ds, Fs, withdrawal) rates.[14,21−23]

Considering flipped classrooms in higher education chemistry, a number of reports have recently appeared, with an early review on emerging trends[9] following a 2015 conference on the topic.[24] These studies have built upon the seminal work of Reid and co-workers,[25,26] who showed that prelecture resources designed to reduce student cognitive load in lecture were effective in eliminating a correlation between initial student qualification and end of term exams in first year university chemistry. Using a case-study approach, Seery and Donnelly in 2012 examined the use of online prelecture resources to enhance a lecture-based course, and reported similar outcomes.[27] Recent longitudinal studies of flipped organic chemistry courses showed in one case a reduction in DFW rate and increase in student grade point average (GPA) for the flipped section,[23] and in another case an improvement in student scores and reduction in withdrawal and failure rates for flipped sections.[28]

In general chemistry, Yestrebsky reported a parallel study (without pretest) of the flipped classroom in a large enrollment class.[29] While there was no reported difference on the

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

standardized final exam scores between the control and flipped sections, an increase in A and B grades, and decrease in C grades, was observed in the flipped section, while the percentage of D and F grades remained the same. No information was provided on the number of withdrawals in the two sections. In 2014, Baepler and co-workers reported a longitudinal study of the flipped class in large enrollment general chemistry courses, using standardized exam scores.[30] Results across the two-year study were mixed, with a statistically significant improvement in the flipped class noted in year 1 but not in year 2. Analysis of scores by quartile using grade point average (GPA) showed an improvement across all demographics in the flipped classroom; however, the differences were not statistically significant. Again, in this study no information was provided on DFW rates. Most recently, Weaver and Sturtevant conducted a longitudinal study of the flipped classroom in a first year course for chemistry majors at Purdue, and reported a statistically significant increase in standardized exam scores for students in the flipped section as compared with the prior lecture-based course.[31]

Beyond effects of the flipped classroom on student performance, important also are student opinions and satisfaction. Here the evidence appears solidly in support of the concept. For example, Smith reported a survey of general chemistry students and found that more than 80% classified the flipped approach as "more useful and/or enlightening", while 65% indicated that pushing lecture content outside the classroom made the class "less boring and/or more engaging".[32] On the other hand, 48% of students agreed that the flipped approach was an extra time burden, and 70% of students were neutral as to the usefulness of the quiz questions in "gauging their own level of understanding" of the material. Other studies have shown high levels of student satisfaction with the flipped class, in comparison with traditional lecture-based courses.[28,33]

In designing our study, we sought to fill gaps in current knowledge by focusing on a nonmajors general chemistry course, noting also that there have been few quantitative, controlled studies of the flipped classroom model over an entire course where lecture and flipped sections taught by the same instructor were run in parallel, rather than longitudinally.[34] Such a study can control for many variables, including instructor, content, delivery, and assessments, yet suffers from a lack of random student selection in the flipped and control sections, which cannot be achieved in a system of open registration. To offset this limitation, a pretest is typically employed.[35] We therefore undertook the following approach. In Fall semester 2014, we piloted the flipped classroom in our off-sequence second-term general chemistry course (69 students). All instructors in the fall first-term course agreed to give the (first-term general chemistry form 2005) American Chemical Society (ACS) exam as a final, which we used as our pretest. Prior to the start of spring registration, we advertised, through classroom announcements and a flyer posted on the departmental Web site and distributed to student advisors across the university, the availability of a flipped section in the spring second-term course, and students self-selected into the spring control and flipped sections, which were taught in parallel by the same instructor. Of the 323 students in the two spring courses, 91% had taken the first-term course in fall, and thus the pretest. Students in the spring courses were given five common exams, culminating in the second-term general chemistry ACS exam (form 2010); we establish the reliability of these instruments using classical test theory and Rasch analysis.[36−39] In this study, we sought to answer the following questions:

(1) What is the effect of the flipped classroom on student performance in a nonmajors general chemistry course, as measured by a series of common exams?
(2) What is the effect of the flipped classroom on DFW rates in a nonmajors general chemistry course?
(3) What is the effect of the flipped classroom on student opinions of the course and instructor?

METHODOLOGY

Population Studied

This study was carried out in two sections of a second-term nonmajors general chemistry course; a separate majors course was available. Table 1 lists the general characteristics of students in the flipped and control (traditional lecture-based) sections. While self-selection could not be avoided, the demographics of the sections are very similar. For example, the sections contained similar percentages of nonmajors (94% flipped, 90% control) and STEM majors (93% in each).

Table 1. Summary of Demographic Information for Control and Flipped Classes

Item	Control		Flipped	
N_{total}	206		117	
N_{study}[a]	187		106	
Gender	49.2% F, 50.8% M		51.9% F, 48.1% M	
Major[b]	Health Science	48.1%	Health Science	39.6%
	Biology	17.6%	Biology	12.3%
	Engineering	16.6%	Engineering	34.0%
	Biochemistry	7.0%	Biochemistry	0.9%
	Arts and Sciences	5.9%	Arts and Sciences	5.7%
	Chemistry	2.7%	Chemistry	4.7%
	Physics	1.6%	Physics	1.9%
	Education	0.5%	Education	0.9%
Pretest Mean (SD)	46.0 (10.7)		46.3 (9.3)	
Pretest Median	47.0		48.5	

[a]Number of students completing pretest (enrolled in CHEM 1001 in prior semester). [b]Arts and Sciences = other majors in A&S than those listed or undeclared.

To account for the nonrandom student selection, we used pretest scores and also student percentile ranking in the fall CHEM 1001 course, taken across all sections. The upper panel of Figure 1 shows the binned pretest scores (raw score, out of 70) for the two courses. The average value was nearly identical (46.1, control; 46.3, FL), with no statistical difference ($p < 0.05$ using one-tailed t-test for independent samples), and the effect size as measured by Cohen's d parameter[40] was 0.002, indicating >99.9% overlap of the two populations. A comparison of the binned percentile ranking, shown in the bottom panel of Figure 1, also reveals little difference, with a slightly higher mean (56% vs 53%) in the control course. Note that the median pretest scores (Table 1) lie significantly above the national norm (39.5).[41]

Course Format

The flipped and control classes were designed to achieve equivalent expectations for student effort, as shown in Figure S1, which gives an outline of the course format and time-

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

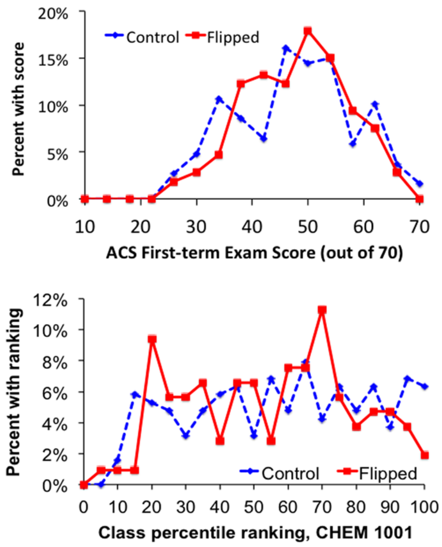

Figure 1. Upper panel: Comparison of pretest score distribution in the flipped and control sections. Lower panel: Comparison of first-term course percentile ranking distribution in the two sections.

expectation comparison for the two courses. The same laboratory curriculum was used in both courses.

The control course featured our standard design: three 50 min lectures each week and a 50 min discussion section (typical enrollment of 35) led by a teaching assistant. The lectures were conducted in a large (300 seat) lecture hall and incorporated clicker questions and demonstrations. Attendance was not required; however, participation points were awarded for lecture and discussion attendance. Weekly online homework sets were assigned using the CONNECT web-based homework system[42]—students were given two chances to complete each assignment. Students were encouraged to post questions to an online discussion board set up on the course management system; however, no credit was given.

In the flipped course, the lecture content was delivered as voice-over Powerpoint based videos, captured and edited using CAMTASIA 2 software[43] with a WACOM Bamboo tablet.[44] The edited video lectures were from 5 to 20 min in length, averaging 13.2 min; these contained the content of the associated classroom lecture. The videos were recorded and edited to move at a fast pace; such videos obviously do not contain time for settling in, transitions, student questions, and other activities that are part of the normal lecture.[9] Three videos were typically assigned each week, made available only on the course management system; each was paired with an assignment delivered using the CONNECT system.[42] To avoid the potential for cross-talk, we did not exactly replicate questions used in the homework set in the control course; rather, we used questions of similar difficulty on the same topic. Table S1 shows a breakdown of the homework/assessment questions by difficulty, based upon classifications in the database.[42] As in the control class, two chances were given to complete each assignment; these were available for credit only within a one-week period. There were no formal lectures, and the discussion sections (typical enrollment of 30) met once each week for 75 min, led by the professor and a teaching assistant. The discussion format was primarily problem solving; however, occasionally demonstrations were performed and short "microlectures" (1−2 slide lectures) given on specific topics by the instructor.[45] Participation points were awarded for discussion attendance and for posting questions to the online discussion board on the course management system (one question or response to a question per week).

The discussion best practices learned in the fall pilot were used in *both* spring courses, in the interest of fairness and student success. Thus, the flipped discussion sets were written on a "just-in-time" basis after reviewing quiz performance and discussion board queries. The same questions, scaled to accommodate the reduced time (50 vs 75 min), were used in the control course. To facilitate peer instruction/interaction, students were placed into discussion groups of three (denoted A1, A2, ...; B1, B2, ...; C1, C2, ...) on the basis of their performance on the first semester ACS exam (one student per group was selected from each third of the distribution). The same discussion groups were kept for the entire semester. Problems were typically written with multiple components, which were then assigned to the different (A, B, C) groups—a design meant to facilitate discussion by examining a problem from multiple aspects. Sample discussion problems are provided in the Supporting Information.

Due to scheduling issues, it was not possible to use the same set of teaching assistants in the two courses. However, student evaluations for discussion TAs in the two sections were comparable.

Instrument Validation

We examined student performance on five common exams, spread over the semester to allow insight into history, maturation, and/or regression of the populations.[35] The validity of these instruments was established using both classical test theory and Rasch analysis.[36−39] The in-semester exams consisted typically of 20 questions; 15 of these were typically multiple choice, the remainder free response. The questions were selected from a test-bank.[46] In establishing the reliability of these assessments, we analyzed a total of 59 multiple choice questions from the four in-semester exams and 70 multiple choice questions from the ACS final exam. All psychometric analysis was performed using JMetrik software.[47]

Figure S2 shows item difficulty−discrimination maps for the in-semester and ACS final exam, as derived from item analysis of 307 student responses (all students in both sections who completed the course). Overall, these plots are very similar, with a similar median discrimination level (0.27) and a slightly larger median difficulty for the in-semester exam questions (0.64 vs 0.55). Instrument reliability estimates in the form of Cronbach's α parameter[48] and Guttman's λ_2 parameter[49] from the item analysis are given in Table S1: these show a smaller, but still acceptable,[50] level of reliability for the in-semester assessments.

Rasch analysis of these items leads to similar conclusions. Here, all questions were scored as correct/incorrect, and the Rasch model for dichotomous data was employed, which expresses the probability of a correct answer by a given person (n) for a given item (i) in terms of parameters expressing ability (β_n) and item difficulty (δ_i):[36−39]

$$P(X_{ni}=1) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \tag{1}$$

These parameters are most conveniently expressed on a logit (=log(odds)) scale displaying both person and item parameters, so-called Wright maps.[51] Figure S3 shows Wright maps generated from separate analysis of the in-semester and ACS exam items for 307 students. These are comparable in item

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

level distribution and show well-centered person measure distributions. Combining all 129 items, the person reliability estimated from Rasch analysis was 0.89, while the item reliability was 0.99 (Table S2).

Student Evaluations

Student evaluation data was collected using the standard University Online Course Evaluation, a 15 item six level Likert scale questionnaire,[52] which also contained a free commentary section. The evaluation submission window was opened at the end of the term, prior to final exams; results were made available after completion of final course grades. The response rate in the two sections was comparable; 57% in the control, 53% in the flipped.

Research Framework

No explicit research framework was employed in this study.

IRB Approval

The study was reviewed and approved (exemption granted) by the Institutional Review Board, and the results reviewed by an academic compliance officer to ensure compliance with all requirements of the Federal Educational Rights and Privacy Act (FERPA).

■ RESULTS

Exam Performance

Our first result is that there was no *overall* difference ($p < 0.05$) in exam performance between the flipped and control sections, neither on the in-semester exams nor on the ACS final; aggregate data are provided in Table S3. The average of all five exams was 61.7 in both classes, with a negligible effect size (Cohen's $d = 0.004$). However, in considering our first research question, does this apply to all student demographics?

The Rasch analysis (Figure S3 and Table S2) indicated three different strata (statistically distinct levels) of students, and therefore we broke down the data by pretest score, grouping students in the control and flipped classes into three bins of comparable size. In the upper panel of Figure 2, we show a comparison of performance (flipped vs control) in terms of the difference in raw score expressed as a percentage, against the binned first-semester ACS exam score. Only in the lowest bin (ACS exam scores ≤ 40) was a consistent trend observed, where a higher score in the flipped class, varying between +5% and +15%, was seen across all five exams. As shown in the lower panel of Figure 2, which compares aggregated results, the level of improvement was similar in this bottom demographic when aggregating exams 1−4 in comparison with the ACS exam, and thus persisted in the overall exam average (8% higher in the flipped section).

Table 2 includes all statistical data related to this comparison, including number of students in each comparison bin, p values, and effect sizes as measured by Cohen's d parameter. Considering again the lowest bin, a statistically significant ($p < 0.05$) difference in overall exam performance between the two sections is observed, with a moderate effect size (0.43).

Figure 3 displays the results of a similar analysis, but with exam scores binned against student percentile ranking in the fall first-term course, taken across all sections. The detailed statistical data are given in Table 3. In this case, improvement across all five exams is seen in both the middle and bottom bins. For the latter, the level of improvement mirrors that seen in Figure 2, and the difference in overall exam average is statistically significant ($p < 0.05$), with a moderate effect size

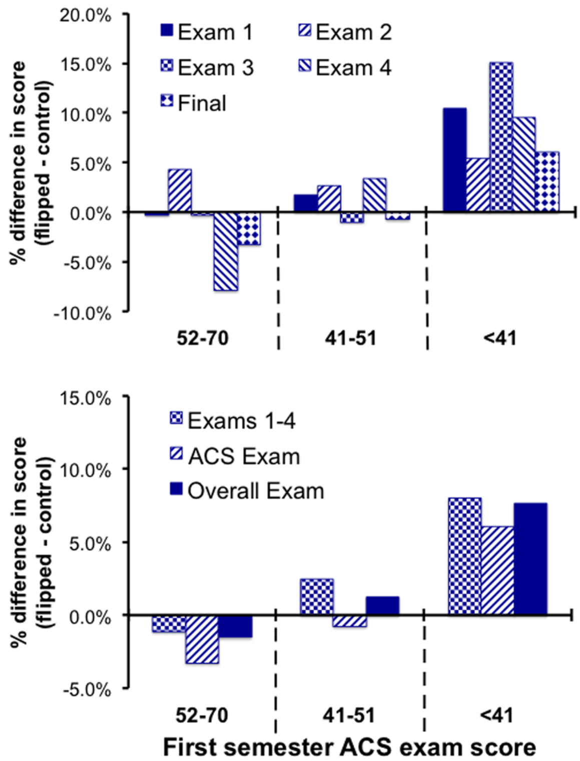

Figure 2. Upper panel: Comparison of individual exam performance in the flipped and control courses, expressed as a percent difference for three different strata of student performance on the pretest. Positive values indicate higher scores in the flipped class. Lower panel: Aggregate exam performance for exams 1−4, compared with performance on the ACS final and the average of all five exams.

(0.39), Table 3. For the middle bin, the difference approaches significance ($p = 0.09$).

Failure and Withdrawal Rates

Addressing our second research question, data concerning the DFW (Ds, Fs, withdrawal) rates (DFW%) in each section are shown following each exam and for the final course assessment in Figure 4. In calculating the DFW rates, we included all course components, using the course grading scale set following Exam 1, which was the same for both sections. For the final assessment, we also compare with results from the flipped off-semester second-term course in Fall 2014, with the third Spring 2015 section (taught in a lecture-based format with a different instructor), and with historical data in the course, which was derived from the previous two years of the on-semester second-term course, encompassing some 1100 students.

These data show a 56% reduction in DFW rate in the flipped course relative to the control, which is remarkably consistent with a recent meta-analysis of active learning in STEM fields, which found a 55% decrease in DFW rates when active learning was employed.[14] Particularly notable was the reduction in withdrawals. Combining the two flipped courses to give a population similar to that in the control, the withdrawal percentage was 1.6% (flipped) vs 6.3% (control); nearly *four times higher* in the latter. In the bottom grade demographic as identified by pretest score, withdrawal rates were 6% in the flipped course, 23% in the control course. As students withdrawing from the course were not included in the analysis presented in Figures 2 and 3, it is probable that the

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 2. Summary of Exam Scores and Analysis, Ranked by Pretest Performance

Item	Parameter	Group 1 (Pretest Score ≤ 40)	Group 2 (Pretest Score 41−51)	Group 3 (Pretest Score 52−70)
Exam 1	$N_{control}$	53	70	61
	Mean (SD)	49.8 (13.8)	65.6 (12.6)	76.4 (12.2)
	$N_{flipped}$	31	43	32
	Mean (SD)	54.9 (13.7)	66.7 (10.6)	76.2 (12.5)
	% diff (f − c)	10.2	1.7	−0.3
	Significance	0.032 (<0.05)	0.44	0.47
	Effect size	0.37	0.09	−0.02
Exam 2	$N_{control}$	53	70	61
	Mean (SD)	47.4 (16.4)	60.0 (10.7)	72.8 (11.5)
	$N_{flipped}$	31	43	32
	Mean (SD)	49.9 (10.3)	61.6 (13.3)	75.9 (12.1)
	% diff (f − c)	5.4	2.7	4.3
	Significance	0.25	0.20	0.12
	Effect size	0.18	0.14	0.26
Exam 3	$N_{control}$	48	70	61
	Mean (SD)	45.4 (12.3)	60.2 (12.6)	76.3 (12.8)
	$N_{flipped}$	31	40	32
	Mean (SD)	52.2 (15.8)	59.7 (12.6)	76.1 (13.2)
	% diff (f − c)	15.1	−1.0	−0.3
	Significance	0.06	0.46	0.47
	Effect size	0.49	0.05	0.02
Exam 4	$N_{control}$	44	70	61
	Mean (SD)	47.6 (13.9)	60.7 (13.2)	76.5 (14.7)
	$N_{flipped}$	29	43	32
	Mean (SD)	52.2 (13.8)	62.7 (12.5)	70.5 (11.9)
	% diff (f − c)	9.5	3.3	−7.9
	Significance	0.23	0.07	0.02 (<0.05)
	Effect size	0.32	0.16	0.43
Final (ACS) exam[a]	$N_{control}$	44	70	61
	Mean (SD)	29.1 (6.8)	36.7 (6.7)	46.3 (6.8)
	$N_{flipped}$	29	43	32
	Mean (SD)	30.8 (5.7)	36.5 (5.8)	44.8 (7.7)
	% diff (f − c)	6.1	−0.8	−3.3
	Significance	0.089	0.33	0.18
	Effect size	0.18	0.14	0.26
Overall exam av	$N_{control}$	44	70	61
	Mean (SD)	47.6 (8.5)	59.8 (7.8)	73.6 (9.3)
	$N_{flipped}$	29	43	32
	Mean (SD)	51.3 (8.2)	60.5 (8.1)	72.5 (9.0)
	% diff (f − c)	7.8	1.3	−1.5
	Significance	0.037 (<0.05)	0.32	0.29
	Effect size	0.43	0.09	0.12

[a]Out of 70.

improvement in scores that we observe among the bottom demographic is underestimated.

Analysis by Gender

Digging deeper into the data, we examined whether there was a difference in performance with respect to gender. Thus, Figure S4 shows the exam performance as binned in Figure 3 (by percentile rank) but broken down further by gender. Interestingly, the overall change in each bin was dominated by one gender; male students in the highest bin, female students in the middle, male students in the lowest bin; in each case the performance of these groups was consistent across all five exams. However, only for male students in the bottom third was the difference in overall exam average statistically significant (p = 0.05). This was also the largest performance differential (10% increase for the flipped class) of any demographic.

Correlations of Overall Exam Performance with Other Indicators

Table 4 provides correlations of overall exam performance for the flipped and control classes against key indicators, including participation, online homework/assessment performance, and lab grade. Data are provided for the correlation of all students and for the bottom demographic (defined as in Figure 2 based on pretest score) using the Pearson product-moment correlation coefficient (r).[40] This coefficient varies from +1 (very strong positive correlation) to −1 (very strong negative correlation). Participation shows a moderate positive association with exam performance ($r \sim 0.4$) for all students, with a stronger correlation for the bottom demographic in the control class. Association of overall exam performance with online homework/assessment score is noticeably higher in the flipped section (r = 0.44) as compared with the control (r = 0.27). The

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

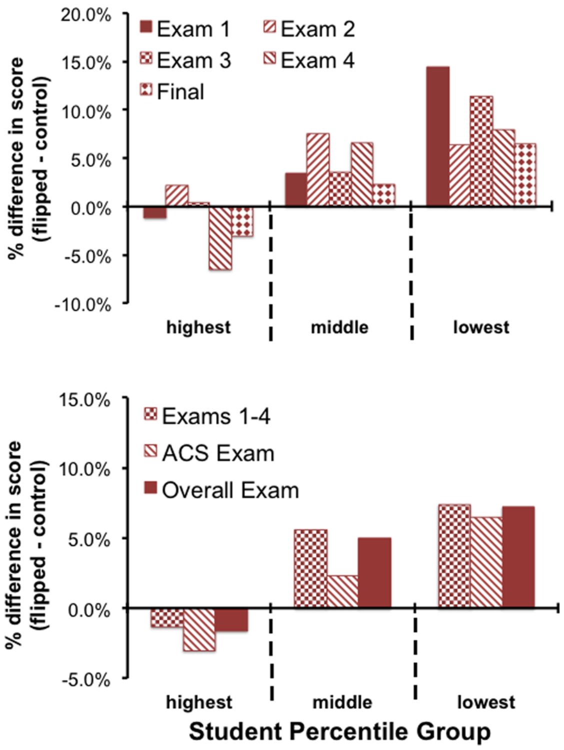

Figure 3. Upper panel: Comparison of individual exam performance in the flipped and control courses, expressed as a percent difference for three different strata of student ranking in the first-term course. Positive values indicate higher scores in the flipped class. Lower panel: Aggregate exam performance for exams 1−4, compared with performance on the ACS final and the average of all five exams.

difference is more pronounced among the bottom demographic, and this disparity increases (r = 0.50, flipped; 0.08, control) if only the first attempt (of 2 allowed) is considered. The correlation with lab grade shows a positive association for all students, weaker for the bottom demographic.

Student Evaluations

Addressing our third research question, student evaluations indicated a high degree of student satisfaction in both spring courses, as shown in Table 5, which provides results of key student evaluation metrics. No statistical difference ($p < 0.05$, one-tailed t-test for independent samples) was found in these evaluation items between the two sections.

More interesting is the comparison of student evaluations in the Fall and Spring *flipped* courses (Table 5). The response rate in the fall course was similar (58%) to that in the spring, and while the design of the flipped courses was identical, student evaluations of the spring course were uniformly higher. For example, responding to the statement "This class positively impacted my problem solving abilities in this subject", the median score increased by 1.1 on the six-point scale, and the difference was statistically significant ($p < 0.001$). Similar gains were observed in the overall course and course content rankings. It was clear from student comments that a minority of students never embraced the flipped format in the fall course. Interestingly, despite the difference in student satisfaction and selection into the course, the DFW% in the two flipped courses was very similar (Figure 4).

■ DISCUSSION

Our results show that the difference in student exam performance in the control and flipped second-term nonmajors general chemistry course is statistically significant only for the bottom third, as measured either by pretest score or percentile ranking in the first-term course. For this demographic, performance is improved in the flipped section, and this trend is mirrored in a lower DFW rate and, particularly, withdrawal percentage. Our findings complement those recently reported in several studies of college-level chemistry courses using different methodologies,[7,28,53] but differ in detail from those previously reported for general chemistry courses,[29,31] which we comment upon further below. While trends are clear from the data, the origin of the trends bears further discussion, and we consider here several possible factors.[18]

We begin with online homework/assessments, the correlation of which with overall exam performance was higher in the flipped class (Table 4), both for the first attempt (of 2) and best attempt. Most interestingly, in the flipped section the correlation with exam performance was similar in the bottom demographic (r = 0.50, first attempt) as for all students (r = 0.52); however, a much weaker correlation was observed in the bottom demographic (r = 0.08) in the control class (Table 4). Table S4 displays statistical data regarding student performance on homework/assessments. The flipped class shows a lower score, on both first and best attempt, for the bottom demographic as compared with all students; however, in the control section, the average for the bottom demographic is slightly *higher* than for all students. Moreover, while the average of best attempt across all students is similar, first attempt scores are significantly lower, on average, in the flipped class.

Data concerning the relative efficacy of online vs written homework is mixed;[54,55] yet, research supports a positive correlation between prelecture homework assignments and higher student achievement.[56−59] However, few studies have examined the effect of online homework *delivery* on student performance. In one study, Raines examined the effect of online homework due dates on exam performance in a college algebra course.[60] For one group, all assignments in a given unit were due just before the unit test; in another, assignment due dates were spaced at intervals between unit tests. No statistically significant difference in exam performance was observed. As questions of similar difficulty on the same topic (Table S1) were used in the online assessments in our study, the primary difference was structural. In the control, a single assessment was available for credit each week. In the flipped section, the weekly assessments were broken down into smaller sets, paired with the content in each video.

The trends in homework/assessment grades (Table S4) may indicate an increased likelihood of unauthorized collaboration (cheating) in the control class. While this is speculative, the issue of cheating in online homework has been discussed in the literature.[61−63] It is hard to track, as normally one cannot be sure of the level of collaboration for a given student. A study in introductory physics at MIT used time per attempt as an indicator of cheating.[62] As we viewed the homework to be formative, there were no time limits placed on individual attempts. Moreover, students were not allowed to directly print the assignments. Thus, time on-assessment recorded for a given student is not necessarily an accurate gauge of the time required for completion. While we were unable to track cheating

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Summary of Exam Scores and Analysis, Sorted by Student Percentile Rank in the First-Term Course

Item	Parameter	Group 1 (Bottom Third)	Group 2 (Middle Third)	Group 3 (Upper Third)
Exam 1	$N_{control}$	44	73	69
	Mean (SD)	48.2 (12.1)	64.6 (12.1)	76.2 (12.1)
	$N_{flipped}$	31	43	32
	Mean (SD)	55.2 (13.3)	66.8 (12.0)	75.3 (11.9)
	% diff (f − c)	14.5	3.5	−1.2
	Significance	0.02 (<0.05)	0.11	0.36
	Effect size	0.51	0.18	−0.07
Exam 2	$N_{control}$	44	73	69
	Mean (SD)	46.3 (16.9)	59.6 (12.0)	71.1 (11.7)
	$N_{flipped}$	31	43	32
	Mean (SD)	49.2 (11.4)	64.2 (11.9)	72.7 (14.7)
	% diff (f − c)	6.4	7.6	2.2
	Significance	0.19	0.03 (<0.05)	0.30
	Effect size	0.20	0.37	0.12
Exam 3	$N_{control}$	39	73	69
	Mean (SD)	46.0 (11.1)	59.1 (15.4)	73.8 (13.6)
	$N_{flipped}$	31	40	32
	Mean (SD)	51.2 (14.9)	61.2 (14.8)	74.2 (12.2)
	% diff (f − c)	11.4	3.6	0.5
	Significance	0.060	0.24	0.45
	Effect size	0.40	0.14	0.03
Exam 4	$N_{control}$	35	73	69
	Mean (SD)	47.9 (14.1)	59.2 (14.9)	74.9 (14.5)
	$N_{flipped}$	29	43	32
	Mean (SD)	51.7 (13.9)	63.1 (13.2)	70.0 (10.5)
	% diff (f − c)	7.9	6.6	−6.5
	Significance	0.15	0.080	0.03 (<0.05)
	Effect size	0.27	0.27	0.36
Final (ACS) exam[a]	$N_{control}$	35	73	69
	Mean (SD)	28.8 (7.1)	36.0 (7.2)	45.3 (7.1)
	$N_{flipped}$	29	43	32
	Mean (SD)	30.7 (5.4)	36.8 (6.6)	44.0 (7.4)
	% diff (f − c)	6.5	2.3	−3.0
	Significance	0.12	0.27	0.19
	Effect size	0.29	0.12	0.19
Overall exam av	$N_{control}$	35	73	69
	Mean (SD)	47.4 (9.2)	58.6 (9.2)	72.1 (9.4)
	$N_{flipped}$	29	43	32
	Mean (SD)	50.8 (7.9)	61.6 (9.3)	71.0 (8.7)
	% diff (f − c)	7.2	5.0	−1.6
	Significance	0.037 (<0.05)	0.089	0.27
	Effect size	0.39	0.32	0.13

[a]Out of 70.

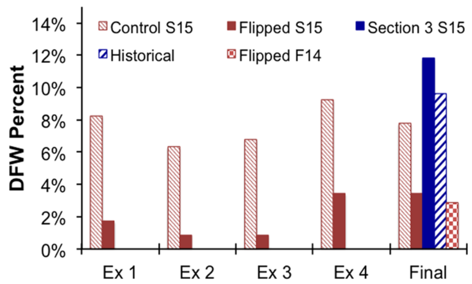

Figure 4. DFW percentage in the control and flipped course, after each in-semester exam (1−4) and the final, calculated as described in the text. Also shown for the final grade are the historical % for the lecture based course, results from the third section in Spring 2015, and results from the flipped classroom pilot in Fall 2014.

directly, our findings suggest that the use of shorter assessments paired with lectures (as in the flipped class) may decrease the likelihood of cheating. This interesting result, which was not a specific research question of this study, bears further study.

Considering other factors, we turn to the degree of student−faculty engagement.[6,8] It is appreciated that large lecture courses can limit student−instructor interaction.[64] Our personal experience highlights the difficulty in engaging the bottom demographic of students in high enrollment courses. In a previous second-term general chemistry course (∼200 students), one of us (S.A.R.) sent individual emails to each student receiving a D or F on the first exam, inviting the student to office hours or by appointment to personally review the exam and standing in the course. Of the 35 students in this group, 2 followed up, some weeks later. In the traditional high enrollment course, a student may easily avoid engaging the

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 4. Correlations of Overall Exam Performance with Key Indicators

Indicator	Comparison Set	Control (Pearson r)	Flipped (Pearson r)
Participation	All students	0.39	0.37
	BD[a]	0.48	0.23
Online homework/ assessment score	All students, best attempt[b]	0.27	0.44
	All students, first attempt[c]	0.30	0.52
	BD, best attempt	0.15	0.42
	BD, first attempt	0.08	0.50
Lab score	All students	0.42	0.47
	BD	0.23	0.34

[a]BD = Bottom Demographic (bottom-third of students as binned by pretest). [b]Best attempt (of 2). [c]First attempt (of 2).

Table 5. Summary of Student Evaluation Items

Item	Key[a]	Control Spring 2015[b]	Flipped Spring 2015	Flipped Fall 2014
How was this class as a whole?	1	5.1 (0.9)	4.9 (0.8)	4.0 (1.4)
How was the content of this class?	1	4.9 (0.8)	4.8 (1.0)	4.2 (1.2)
This class positively impacted my problem solving abilities in this subject.	2	5.3 (0.8)	5.4 (0.8)	4.3 (1.2)
This class was intellectually challenging.	2	5.8 (0.7)	5.7 (0.8)	5.6 (0.8)
Assistance or extra help were available outside of class time.	2	5.8 (0.7)	5.7 (0.9)	5.5 (0.9)
Expectations of students were presented clearly.	2	5.6 (0.8)	5.4 (0.8)	4.8 (1.1)
The instructor provided explanations that reduced confusion.	2	5.7 (0.7)	5.7 (0.5)	5.3 (1.0)
The instructor was well organized.	2	5.9 (0.4)	5.8 (0.5)	5.5 (0.6)
The instructor encouraged student participation.	2	5.6 (0.8)	5.6 (0.6)	5.6 (0.7)

[a]Key 1: 6 = excellent, 5 = very good, 4 = good, 3 = fair, 2 = poor, 1 = very poor. Key 2: 6 = strongly agree, 5 = agree, 4 = agree somewhat, 3 = neither agree nor disagree, 2 = disagree somewhat, 1 = strongly disagree. [b]The group median response and (in parentheses) standard deviation for each item are shown.

professor on a one-to-one level. This is harder to avoid in the flipped course, where the instructor's role shifts to working side-by-side with small groups of students.

Another consideration is student engagement with the course. One metric for this is lecture/discussion attendance. In both sections, participation points were awarded for attendance, and the attendance in the control course (83% lecture, 86% discussion) was essentially identical to that in the flipped (83%). The difference was also small in the bottom demographic, as defined by pretest. There is a positive association (Table 4) of overall exam performance with participation when considering all students (r = 0.39 for control, 0.37 for flipped). Among the bottom demographic as defined by the pretest, the association is slightly stronger in the control class (r = 0.48), and weaker in the flipped.

A related consideration concerns lecture habits, as students in large lecture courses often assume a passive role, as receivers of information (i.e., note-takers or observers).[65] More effective note-taking is correlated with higher exam performance,[66] and, yet, estimates are that first-year students record only 10% of the important points in a typical lecture.[67] Here, pacing is important. In the flipped class, on-demand access to lecture content allows lectures to be viewed at one's own choosing and pace. Most students in the flipped section commented positively on this flexibility, which also ties into student attention span. Recent research suggests that the concept of a single attention span is not well founded,[68] yet attention spans of the first-year student are much shorter than the lecture period (50 min), and likely decrease as the lecture progresses.[69] Moreover, the physical location of a student in the classroom is a factor, due to variations in distractors.[70] In their study of large lecture classes, Benedict and Hoag found that students sitting near the rear of the class had a 25% higher probability of receiving a D or F.[71] Various strategies including clicker questions and in-class demonstrations were used to periodically "re-engage" students in the lecture course; however, the video lectures, which were faster paced and averaged around 13 min in length, are intrinsically better matched to student attention spans.[6,8]

Summarizing, important factors that may lead to the trend we observe in a higher exam performance in the flipped course among the lowest demographic include (1) a stronger correlation of homework/assessment score in the flipped class, which may reflect the effect of pairing online assessments with video lectures; (2) increased student–instructor interaction; (3) mitigation of poor note-taking through the availability of on-demand, self-paced lectures; and (4) a more effective engagement with the lecture material through the ability to select when that engagement occurs and better match of lecture length to student attention span. These factors may also help explain the lack of a statistically significant ($p < 0.05$) difference between the flipped and control courses in the higher student demographics, as in our experience these students usually sit near the front, are more engaged in the lecture, attend instructor office hours, and have better note-taking skills. Moreover, it should be emphasized that, fundamentally, our comparison does not set an active classroom against a passive one, but involves classroom structures that each bear active learning components (e.g., online assessments, discussion sections).

Our findings can be compared with the recent longitudinal study of a first-term majors chemistry class, where improvement in standardized exam performance was observed across every grade demographic.[31] Our results are not necessarily at odds, as we observe a statistically significant difference in only a single demographic. Beyond the differences in design (longitudinal vs parallel, single instructor vs multiple instructor) and population (majors vs nonmajors, small enrollment vs large), an important distinction involves the use of online homework. In Weaver's study, online homework was not used in the traditional course, although written weekly assignments were a component of both courses.[31] The presence of a "quizzing effect" was considered as a possible limitation; however, it was suggest that the quizzes served more as a pacing mechanism, designed to motivate students to watch the lecture.[31] The effect of quizzing was examined in a recent parallel study of graduate student performance in physiology,[34] where the significant overall improvement (>12%) in exam score for students in the flipped section correlated strongly with quiz score (r = 0.77). However, as in Weaver's study, online assessments were not used in the control course. We believe that this is an important distinction, as we found in both

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

sections a positive correlation of exam performance with online assessments, albeit stronger in the flipped course (Table 4).

Weaver suggested one model for application of the flipped approach to a large general chemistry lecture, which involved large discussion sections (of order 100 students) meeting on a twice monthly pattern.[31] We are also interested in scaling up the discussion size; however, based upon the experience gained in this study, we believe that a weekly meeting pattern is optimal, at least in a nonmajors course.

■ CONCLUSIONS

We report a parallel controlled study of the flipped classroom in a second-term general chemistry course. Student populations in the two courses were similar demographically and statistically as measured by pretest and percentile ranking in the first-term course. The courses were designed to require equivalent student effort, and five common exams were administered, culminating in the second-term ACS general chemistry exam (form 2010). Validity of these instruments was calibrated using classical test theory and Rasch analysis, which indicated three strata of student performance. Overall, the exam performance was statistically identical in the two sections; however, when separating into different strata, a statistically significant increase in cumulative exam performance was observed in the bottom-third of students, as defined by pretest or class percentile rank. Following this trend, a significant (56%) decrease in DFW% in the flipped course was found as compared with the control.

Using course data and student evaluations, we examined several factors that may lead to higher exam scores in the flipped course among the lowest demographic, and a reduced DFW rate. The stronger correlation of homework/assessment score with overall course grade is one factor, which reflects a potential advantage in the flipped class design, which paired shorter online assessments with each video lecture. Also potentially important are (1) increased student–instructor interaction in the flipped class, (2) mitigation of poor note-taking through the availability of on-demand, self-paced lectures; and (3) a more effective engagement with the lecture material through the ability to select when that engagement occurs and a better match of lecture length to student attention span. Additional studies will be needed to pinpoint which of these factors is most important.

Finally, while both spring courses showed a high level of student satisfaction, it is clear that student satisfaction in the flipped course increased when the course model was publicized prior to registration and self-selection was allowed.

■ LIMITATIONS

While the parallel experimental design offers many advantages, it also has drawbacks. First and foremost is the lack of true randomness in the student populations, which cannot be achieved in practice, at least in a system of open registration. We sought to overcome this through the use of a pretest, and by considering student rank in the first-term course. The consistency of results for these two benchmarks (compare, e.g., Figures 2 and 3) gives added confidence. However, while equal opportunity was provided for enrollment in the flipped section, the demographic profile of students in the two sections was similar (Table 1), and the pretest scores indicated statistically identical populations (Figure 1), one cannot easily gauge other variables, such as student motivation, which may affect the validity of our control. Longitudinal studies are not immune to this effect either, as student motivations may significantly change over the course of the first year. Moreover, the underlying reasons motivating a student to take a particular section are many, ranging from flexibility of scheduling to choice of instructor to format and more. In this vein, we note that one of our key findings, i.e., a reduction in withdrawal rate, was consistent also with the fall flipped course (Figure 4), where only one section was available, and students were not informed of the course model prior to registration.

A second drawback of the parallel design is the potential for cross-talk between the sections, most problematically, the sharing of resources. We addressed this by restricting access to the video content in the flipped course to the course management system, and by informing students in the syllabus that the videos and related content were the intellectual property of the instructor and could not be shared with anyone outside of the course without his permission. However, we did not prevent downloading of the videos, and the possibility for sharing of resources remained. We made no attempt to assess how widespread this was. However, as the content of the video and in-class lectures was identical, it is not clear what advantage this would afford.

The instructor in the course was aware of the ACS exam content, having given the same exam in the off-semester pilot; however, we do not consider as significant the influence of teaching to the posttest. The content for both courses was developed in the summer prior to the fall pilot, and the only significant modification to the curriculum in the interim between fall and spring semesters was a redesign of one module (organic chemistry). The two sections used the same content and discussion problems, and featured homework/assessments that utilized like questions of similar difficulty. Moreover, the level of improvement exhibited by the bottom third in the in-semester assessments well matched that observed in the posttest, irrespective of whether students were binned by pretest score or first-term percentile rank.

■ IMPLICATIONS

Our results are consistent with the idea that active learning holds particular benefits for students who are capable but less well prepared.[13] As far as these results can be generalized to the first year general chemistry curriculum, it has important implications. Chemistry courses have some of the highest first year DFW rates, and thus flipped courses may be one strategy for increased student success. This should be even more important in the first-term, where DFW rates are traditionally higher. Moving forward, we plan to continue and expand flipped classroom sections of our general chemistry courses, letting student (and instructor) preference be our guide. Of particular interest in the next phase is to scale up the flipped classroom by increasing discussion size. This requires specific infrastructure (e.g, spaces to hold discussions in the round for 50+ students, with white/smart boards), and we hope that this study will motivate the development of improved learning spaces for flipped courses.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00717.

Figures, tables, and discussion set (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: scott.reid@marquette.edu.

*E-mail: michael.ryan@marquette.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Gabriela Weaver, whose work on the flipped classroom motivated this study, and the teaching assistants in both sections (Disha Gandhi, Jeewantha Hewage, Elena Ivanova, Alexander Teplukin, Remigio Usai, Brian Pattengale, Yu Tang). Professors Anita Manogaran and Adam Fiedler provided useful input, and we thank Prof. Rajendra Rathore for helpful guidance on the structure of the organic chemistry module. Finally, we dedicate this article to the late Dr. Elena Ivanova, whose light was too soon extinguished but will never fade from our memory.

REFERENCES

(1) Baker, J. W. In *Selected Conference Papers, Proceedings of the 11th International Conference on College Teaching and Learning*, Jacksonville, FL, 2000; p 9.

(2) Lage, M. J.; Platt, G. J.; Treglia, M. Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *J. Econ. Educ.* **2000**, *31*, 30.

(3) Lage, M. J.; Platt, G. The Internet and the Inverted Classroom. *J. Econ. Educ.* **2000**, *31*, 11.

(4) Tucker, B. The Flipped Classroom. *Educ. Next* **2012**, *12*, 82.

(5) *Implementation and Critical Assessment of the Flipped Classroom Experience*; Scheg, A. G., Ed.; Information Science Reference: Hershey, PA, 2015.

(6) Bergmann, J.; Sams, A. *Flip Your Classroom: Reach Every Student in Every Class Every Day*; International Society for Technology in Education: Eugene, OR, 2012.

(7) Christiansen, M. A. Inverted Teaching: Applying a New Pedagogy to a University Organic Chemistry Class. *J. Chem. Educ.* **2014**, *91*, 1845.

(8) Bergmann, J.; Sams, A. *Flipped Learning: Gateway to Student Engagement*; International Society for Technology in Education: Eugene, OR, 2014.

(9) Seery, M. K. Flipped learning in higher education chemistry: emerging trends and potential directions. *Chem. Educ. Res. Pract.* **2015**, *16*, 758.

(10) Hake, R. R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* **1998**, *66*, 64.

(11) Michael, J. Where's the evidence that active learning works? *Adv. Physiol Educ* **2006**, *30*, 159.

(12) Wenderoth, M. P.; Freeman, S.; O'Connor, E. Prescribed active learning increases performance in introductory biology. *FASEB J.* **2007**, *21*, A220.

(13) Haak, D. C.; HilleRisLambers, J.; Pitre, E.; Freeman, S. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science* **2011**, *332*, 1213.

(14) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8410.

(15) Deslauriers, L.; Schelew, E.; Wieman, C. Improved Learning in a Large-Enrollment Physics Class. *Science* **2011**, *332*, 862.

(16) O'Dowd, D. K.; Aguilar-Roca, N. Garage Demos: Using Physical Models to Illustrate Dynamic Aspects of Microscopic Biological Processes. *Cbe-Life Sci. Educ* **2009**, *8*, 118.

(17) Schultz, D.; Duffield, S.; Rasmussen, S. C.; Wageman, J. Effects of the Flipped Classroom Model on Student Peformance for Advanced Placement High School Chemistry Students. *J. Chem. Educ.* **2014**, *91*, 1334.

(18) Goodwin, B.; Miller, K. Evidence on Flipped Classrooms Is Still Coming In. *Educ. Leadership* **2013**, *70*, 78.

(19) Bishop, J. L.; Verleger, M. A. In *120th ASEE Annual Conference and Exposition*, Atlanta, GA, 2013.

(20) O'Flaherty, J.; Phillips, C. The use of flipped classrooms in higher education: A scoping review. *Internet Higher Educ.* **2015**, *25*, 85.

(21) Bidwell, A. Flipped Classroom May Help Weaker STEM Students. *U.S. News and World Report* [Online Early Access]. Published Online: 2014. http://www.usnews.com/news/stem-solutions/articles/2014/08/05/taking-a-page-from-humanities-college-engineering-gets-flipped (accessed August 5, 2015).

(22) Long, K. Washington college instructors are "flipping" the way they teach. *The Seattle Times* [Online Early Access]. Published Online: 2012. http://www.seattletimes.com/seattle-news/washington-college-instructors-are-flipping-the-way-they-teach/ (accessed August 1, 2015).

(23) Fautch, J. M. The flipped classroom for teaching organic chemistry in small classes: is it effective? *Chem. Educ. Res. Pract.* **2015**, *16*, 179.

(24) American Chemical Society Division of Chemical Education Committee on Computers in Chemical Education. *2014 Spring ConfChem: Flipped Classroom*, 2015. http://confchem.ccce.divched.org/2014SpringConfChem (accessed August 15, 2015).

(25) Sirhan, G.; Gray, C.; Johnstone, A. H.; Reid, N. Preparing the Mind of the Learner. *Univ. Chem. Educ.* **1999**, *3*, 43.

(26) Sirhan, G.; Reid, N. Preparing the Mind of the Learner - Part 2. *Univ. Chem. Educ.* **2001**, *5*, 52.

(27) Seery, M. K.; Donnelly, R. The implementation of pre-lecture resources to reduce in-class cognitive load: A case study for higher education chemistry. *Brit J. Educ Technol.* **2012**, *43*, 667.

(28) Flynn, A. B. Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French. *Chem. Educ. Res. Pract.* **2015**, *16*, 198.

(29) Yestrebsky, C. *Flipping the Classroom in a Large Chemistry Class-research University Environment Procedia - Social and Behavioural Sciences* **2015**, *191*, 1113.

(30) Baepler, P.; Walker, J. D.; Driessen, M. It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Comput. Educ* **2014**, *78*, 227.

(31) Weaver, G. C.; Sturtevant, H. G. Design, Implementation, and Evaluation of a Flipped Format General Chemistry Course. *J. Chem. Educ.* **2015**, *92*, 1437.

(32) Smith, J. D. Student attitudes toward flipping the general chemistry classroom. *Chem. Educ. Res. Pract.* **2013**, *14*, 607.

(33) Yeung, K.; O'Malley, P. J. Making 'the flip' work: barriers to and implementation strategies for introducing flipped teaching methods into traditional higher education courses. *New Directions* **2014**, *10*, 59.

(34) Tune, J. D.; Sturek, M.; Basile, D. P. Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Adv. Physiol Educ* **2013**, *37*, 316.

(35) Creswell, J. W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 4th ed.; Pearson: Boston, 2012.

(36) Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Danish Institute for Educational Research: Copenhagen, 1960.

(37) Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests (Expanded ed.)*; The University of Chicago Press: Chicago, 1980.

(38) Boone, W. J.; Scantlebury, K. The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Sci. Educ.* **2006**, *90*, 253.

(39) Scantlebury, K.; Boone, W. J. Designing Tests and Surveys for Chemistry Education Research. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series *976*; American Chemical Society: Washington, DC, 2008; p 14910.1021/bk-2008-0976.ch010.

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX

(40) Cohen, J. *Statistical power analysis for the behavioral sciences*, 2nd ed.; Erlbaum Associates: Hillsdale, NJ, 1988.
(41) ACS Exams Institute. http://chemexams.chem.iastate.edu/instructors/exam-statistics/national-norms (accessed August 21, 2015).
(42) CONNECT McGraw-Hill, 2015, http://connect.customer.mheducation.com/about/.
(43) Camtasia for Mac, Version 2.8.3, TechSmith Corporation, 2014, http://www.techsmith.com/camtasia.html.
(44) Wacom Bamboo Tablet, 2014. http://www.wacom.com/en-us/products/navigation/bamboo-pad-usb.
(45) Sweet, D. Microlectures in a Flipped Classroom: Application, Creation and Resources. *Mid-West. Educ. Res.* **2014**, *26*, 52.
(46) Chang, R.; Goldsby, K. A. *Chemistry*, 11th ed. Testbank; McGraw-Hill: 2014.
(47) JMetrik, Version 3.1.2, 2014, http://www.jmetrik.com.
(48) Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* **1951**, *16*, 297.
(49) Guttman, L. A Basis for Analyzing Test-Retest Reliability. *Psychometrika* **1945**, *10*, 255.
(50) George, D.; Mallery, P. *SPSS for Windows step by step: A simple guide and reference (11.0 update)*, 4th ed.; Allyn and Bacon: Boston, 2003.
(51) Wilson, M. On Choosing a Model for Measuring. *Methods Psychol. Res.-Online* **2003**, *8*, 1.
(52) Sample MOCES Instrument. http://www.marquette.edu/oira/cevaldocuments/evaluation_form.pdf (accessed August 20, 2015).
(53) Trogden, B. G. In *Implementation and Critical Assessment of the Flipped Classroom Experience*; Scheg, A. G., Ed.; Information Science Reference: Hershey, PA, 2015.
(54) Fynewever, H. A comparison of the effectiveness of web-based and paper based homework for General Chemistry. *Chem. Educ.* **2008**, *13*, 264.
(55) Arasasingham, R. D.; Taagepera, M.; Potter, F.; Martorell, I.; Lonjers, S. Assessing the effect of Web-based learning tools on student understanding of stoichiometry - Using knowledge space theory. *J. Chem. Educ.* **2005**, *82*, 1251.
(56) Freasier, B.; Collins, G.; Newitt, P. A web-based interactive homework quiz and tutorial package to motivate undergraduate chemistry students and improve learning. *J. Chem. Educ.* **2003**, *80*, 1344.
(57) Botch, B.; Day, R.; Vining, W.; Stewart, B.; Rath, K.; Peterfreund, A.; Hart, D. Effects on student achievement in general chemistry following participation in an online preparatory course - ChemPrep, a voluntary, self-paced, online introduction to chemistry. *J. Chem. Educ.* **2007**, *84*, 547.
(58) Richards-Babb, M.; Jackson, J. K. Gendered responses to online homework use in general chemistry. *Chem. Educ. Res. Pract.* **2011**, *12*, 409.
(59) Eichler, J. F.; Peeples, J. Online Homework Put to the Test: A Report on the Impact of Two Online Learning Systems on Student Performance in General Chemistry. *J. Chem. Educ.* **2013**, *90*, 1137.
(60) Raines, J. M. The Effect of Online Homework Due Dates on College Student Achievement in College Algebra. *J. Stud. Educ.* **2012**, DOI: 10.5296/jse.v2i3.1704.
(61) Stephens, J. M.; Young, M. F.; Calabrese, T. Does moral judgment go offline when students are online? A comparative analysis of undergraduates' beliefs and behaviors related to conventional and digital cheating. *Ethics Behav* **2007**, *17*, 233.
(62) Palazzo, D. J.; Lee, Y. J.; Warnakulasooriya, R.; Pritchard, D. E. Patterns, correlates, and reduction of homework copying. *Phys. Rev. ST Phys. Educ. Res.* **2010**, *6*, No. 010104.
(63) Hamlen, K. R. Academic dishonesty and video game play: Is new media use changing conceptions of cheating? *Comput. Educ* **2012**, *59*, 1145.
(64) Wulff, D. H.; Nyquist, J. D.; Abbott, R. D. In *New directions for teaching and learning: teaching large classes well*; Weimer, M., Ed.; Jossey-Bass: San Francisco, CA, 1987; p 17.
(65) McKeachie, W. J. *Teaching tips: strategies, research and theory for college and university teachers*; 10th ed.; Houghton Mifflin: New York, 1999.
(66) Johnstone, A. H.; Su, W. Y. Lectures - a learning experience? *Educ. Chem.* **1994**, *31*, 75.
(67) Locke, E. A. An empirical study of lecture notetaking among college students. *Journal of Educational Research* **1977**, *77*, 93.
(68) Wilson, K.; Korn, J. Attention During Lectures: Beyond Ten Minutes. *Teaching of Psychology* **2007**, *34*, 85.
(69) Bunce, D. M.; Flens, E. A.; Neiles, K. Y. How long can students pay attention in class? A study of student attention decline using clickers. *J. Chem. Educ.* **2010**, *87*, 1438.
(70) Holliman, W. B.; Anderson, H. N. Proximity and student density as ecological variables in a college classroom. *Teaching of Psychology* **1986**, *13*, 200.
(71) Benedict, M. E.; Hoag, J. Seating location in large lectures: Are seating preferences or location related to course performance? *Journal of Economic Education* **2004**, *35*, 215.

DOI: 10.1021/acs.jchemed.5b00717
J. Chem. Educ. XXXX, XXX, XXX−XXX